ÖREBRO
UNIVERSITET

Lösning till tentamen i

Matematisk statistik och sannolikhetslära

MA506G

2021-06-04

---

1. Fördelningsfunktionen för den kontinuerliga slumpvariabeln X ges av $F_X(x) = (3x - x^3)/2$ för $0 \leq x \leq 1$.

   (a) Beräkna $\mathbf{P}(0.25 \leq X \leq 0.5)$. [3p]

   (b) Beräkna $\mathbf{E}(X)$ och $\mathbf{V}(X)$. [4p]

   (c) Är medianen större eller mindre än 0.4? Glöm inte att motivera väl. [3p]

   Lösning:

   (a) Vi har

$$\begin{aligned}\mathbf{P}(0.25 \leq X \leq 0.5) &= F_X(0.5) - F_X(0.25)\\ &= \frac{3 \cdot 0.5 - 0.5^3 - 3 \cdot 0.25 + 0.25^3}{2}\\ &= \frac{\frac{3}{4} - \frac{7}{64}}{2} = \frac{41}{128} = 0.32.\end{aligned}$$

   (b) Täthetsfunktionen ges av $f_X(x) = F_X'(x) = 3(1 - x^2)/2$. Därmed får vi

$$\begin{aligned}\mathbf{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x\\ &= \int_0^1 \frac{3}{2}\left(x - x^3\right)\,\mathrm{d}x\\ &= \frac{3}{2}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_0^1 = \frac{3}{8},\\ \mathbf{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)\,\mathrm{d}x\\ &= \int_0^1 \frac{3}{2}\left(x^2 - x^4\right)\,\mathrm{d}x\\ &= \frac{3}{2}\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_0^1 = \frac{1}{5},\end{aligned}$$

och
$$\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \frac{1}{5} - \frac{9}{64} = \frac{19}{320} = 0.0594.$$

(c) Eftersom $F_X(0.4) = 0.5680 > 0.5$ är mer än hälften av sannolikheten samlad under 0.4. Därmed är medianen också lägre än 0.4.

2. I volleyboll spelar man först till tre set. Varje set kan ses som ett oberoende försök med vinstsannolikhet p. När Örebro Volley spelar mot Gislaved är sannolikheten att Örebro vinner setet $p = 0.6$.

   (a) Bestäm sannolikhetsfunktionen för Örebro segermarginal, som är antalet set som Örebro vinner minus antalet set som Gislaved vinner (om Örebro förlorar är alltså marginalen negativ). [4p]

   (b) Vid resultaten 3–0 och 3–1 tilldelas vinnande lag 3 poäng och förlorande lag 0 poäng. Vid resultatet 3–2 tilldelas vinnande lag 2 poäng och förlorande lag 1 poäng. Beräkna väntevärdet av Örebros poängsumma? [2p]

   (c) Bestäm sannolikheten att Örebro vann första set om de tilldelats 3 poäng. [4p]

   Lösning:

   (a) Slumpvariabeln Z kan anta värdena {-3, -2, -1, 1, 2, 3}. Låt $X \in \text{NegBin}(3, 0.6)$. För positiva värden inser vi att
$$p_Z(n) = p_X(6-n) = \binom{5-n}{2} \cdot 0.6^3 \cdot 0.4^{3-n},$$
   så vi har
$$\begin{aligned} p_Z(3) &= \binom{2}{2} \cdot 0.6^3 \cdot 0.4^0 = 0.216 \\ p_Z(2) &= \binom{3}{2} \cdot 0.6^3 \cdot 0.4^1 = 0.2592 \\ p_Z(1) &= \binom{4}{2} \cdot 0.6^3 \cdot 0.4^2 = 0.20736 \end{aligned}$$
   För negativa värden kan vi byta ut p mot $1-p$. Med $Y \in \text{NegBin}(3, 0.4)$ får vi då
$$p_Z(-n) = p_Y(6-n) = \binom{5-n}{2} \cdot 0.4^3 \cdot 0.6^{3-n},$$

vilket ger

$$p_Z(-1) = \binom{4}{2} \cdot 0.4^3 \cdot 0.6^2 = 0.13824$$
$$p_Z(-2) = \binom{3}{2} \cdot 0.4^3 \cdot 0.6^1 = 0.1152$$
$$p_Z(-3) = \binom{2}{2} \cdot 0.4^3 \cdot 0.6^0 = 0.064$$

(b) Låt W vara poängsumman. Vi har

$$p_W(3) = p_Z(3) + p_Z(2) = 0.4752$$
$$p_W(2) = p_Z(1) = 0.20736$$
$$p_W(1) = p_Z(-1) = 0.13824$$
$$p_W(0) = p_Z(-2) + p_Z(-3) = 0.1792$$

Detta ger

$$\mathbf{E}(W) = \sum_k k p_W(k) = 1.97856.$$

(c) Låt B vara händelsen att Örebro vinner första set och A händelsen att de tilldelas 3 poäng. Enligt Bayes sats har vi

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(B)\mathbf{P}(A|B)}{\mathbf{P}(A)}.$$

Vi har $\mathbf{P}(B) = p$ och $\mathbf{P}(A) = 0.216 + 0.2592 = 0.4752$. Vidare ges, på samma sätt som ovan

$$\mathbf{P}(A|B) = \binom{1}{1} \cdot 0{,}6^2 \cdot 0{,}4^0 + \binom{2}{1} \cdot 0{,}6^2 \cdot 0{,}4^1 = 0.36 + 0.288 = 0.648.$$

Vi får därmed

$$\mathbf{P}(B|A) = \frac{0.6 \cdot 0.648}{0.4752} = 0.8182.$$

3. I ett nytt bostadsområde med 1000 bostäder planeras för nya grundskolor. Antalet barn per familj antas vara oberoende, och sannolikheterna för antalet grundskolepliktiga barn i en familj skattas för 0, 1, 2 och 3 barn till 0.4, 0.2, 0.3 respektive 0.1. Hur stor är sannolikheten att 1150 grundskoleplatser kommer att räcka till samtliga barn? [10p]

Lösning: Låt X_j vara slumpvariabeln som anger antalet grundskolepliktiga barn i familj j. Vi kan approximera $X = \sum X_j$ med en normalfördelad slumpvariabel enligt centrala gränsvärdessatsen, tack vare oberoende.

Vi har

$$\mathbf{E}(X_1) = 0 \cdot 0.4 + 1 \cdot 0.2 + 2 \cdot 0.3 + 3 \cdot 0.1 = 1.1,$$
$$\mathbf{E}(X_1^2) = 0 \cdot 0.4 + 1 \cdot 0.2 + 4 \cdot 0.3 + 9 \cdot 0.1 = 2.3$$

och

$$\mathbf{V}(X_1) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = 2.3 - 1.1^2 = 1.09.$$

Detta ger $\mathbf{E}(X) = 1000\mathbf{E}(X_1) = 1100$ och $\mathbf{V}(X) = 1000\mathbf{V}(X_1) = 1090$.

Därmed kan X approximeras med $Y \in N(1100, 1090)$. Vi får

$$\begin{aligned}\mathbf{P}(X \leq 1150) &\approx \mathbf{P}(Y \leq 1150.5)\\ &= \mathbf{P}\left(\frac{Y - 1100}{\sqrt{1090}} \leq \frac{1150.5 - 1100}{\sqrt{1090}}\right)\\ &= \Phi(50.5/33.0151) = \Phi(1.5296) = 0.937.\end{aligned}$$

Sannolikheten är alltså 93.7%.

4. Elasticiteten i en typ av konstfiber förväntas öka om man tillsätter flour. Vi gör ett antal observationer med olika halter av flour tillsatt, och får då följande data: [10p]

Flourhalt	0.0	0.3	0.7	1.0	1.5	2.0	2.5	3.0
Elasticitet	12.86	13.10	13.82	14.60	15.72	16.47	16.13	18.45

Anpassa en linje till dessa punkter med linjär regression, och rita sedan ut såväl punkter som linje i ett spridningsdiagram. Beräkna slutligen förklaringsgraden.

Lösning: Vi kallar flourhalterna för $x_1, \ldots, x_8$ och elasticiteterna för $y_1, \ldots, y_8$. Deras medelvärden blir

$$\bar{x} = \frac{x_1 + \cdots + x_8}{8} = 1.375$$

och

$$\bar{y} = \frac{y_1 + \cdots + y_8}{8} = 15.14.$$

Vi får då

$$\begin{aligned}S_{xx} &= \sum_k (x_k - \bar{x})^2 = 7.955\\ S_{yy} &= \sum_k (y_k - \bar{y})^2 = 25.44\\ S_{xy} &= \sum_k (x_k - \bar{x})(y_k - \bar{y}) = 13.82,\end{aligned}$$

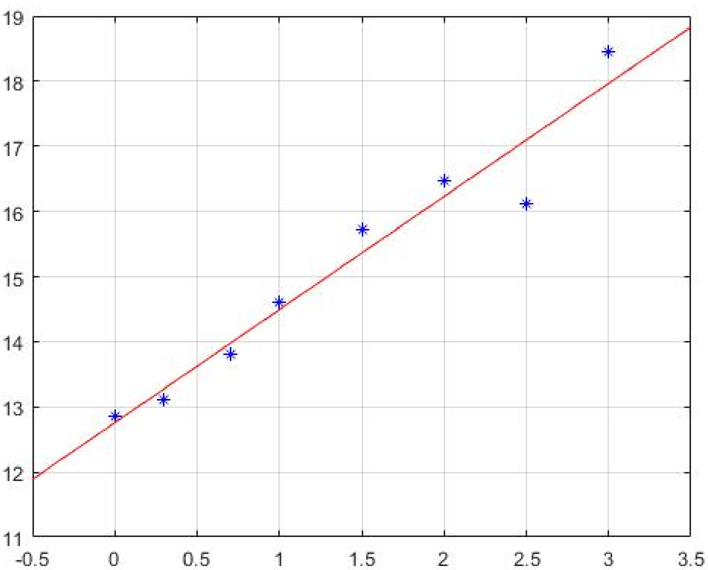
vilket ger

$$\beta^* = \frac{S_{xy}}{S_{xx}} = \frac{13.82}{7.955} = 1.737$$

och

$$\alpha^* = \bar{y} - \beta^* \bar{x} = 15.14 - 1.737 \cdot 1.375 = 12.75.$$

Slutligen ritar vi ut våra värden och vår regressionslinje:



5. (a) En drönare får sin strömförsörjning antingen från två batterier med spänning 7.2 volt eller fyra batterier med spänning 3.6 volt. Sannolikheten att batteriet slutar att fungera inom en timme är p, oavsett batterityp, och batterierna förutsätts fungera oberoende av varandra. För att drönaren inte ska störta behöver minst hälften av batterierna fungera. För vilka p är det säkrare att flyga en timme med två batterier än fyra? [6p]

(b) En annan drönarmodell har utrustats med 125 minibatterier, som vardera har sannolikheten $p = 0.08$ att fungera en hel timme. Hur stor är sannolikheten att drönaren kan flyga en timme, om högst 12 batterier kan avvaras? [4p]

Lösning:

(a) Antalet trasiga batterier är binomialfördelat, antingen $X \in \text{Bin}(2, p)$ eller $Y \in \text{Bin}(4, p)$. Vi har

$$\mathbf{P}(X = 2) = p^2$$

och

$$\mathbf{P}(Y \geq 3) = 4p^3(1-p) + p^4 = 4p^3 - 3p^4.$$

Två batterier är säkrare än fyra om $\mathbf{P}(X = 2) < \mathbf{P}(Y \geq 3)$, det vill säga om

$$0 < 4p^3 - 3p^4 - p^2 = p^2(3p-1)(1-p),$$

vilket är uppfyllt om $1/3 < p < 1$.

(b) Slumpvariabeln $Z \in \text{Bin}(125, 0.08)$ kan approximeras med $W \in \text{Po}(125 \cdot 0.08) = \text{Po}(10)$, eftersom $p < 0.1$. Vi får

$$\mathbf{P}(Z \leq 12) \approx \mathbf{P}(W \leq 12) = 0.7916$$

enligt tabell.

6. Ur ett större parti sockerbetor väljs 30 av dessa ut för mätning av sockerhalten, som kan antas vara normalfördelad med väntevärde μ och standardavvikelse σ. Mätningarna ger vid handen att medelvärdet av sockerhalten är $\bar{x} = 16.51$ och att

$$\sum_{k=1}^{30} (x_k - \bar{x})^2 = 45.63.$$

(a) Bestäm ett tvåsidigt 99 % konfidensintervall för μ och ett ensidigt 95 % konfidensintervall för σ som ligger så lågt som möjligt. [6p]

(b) Hur stort stickprov behöver man ta om bredden på konfidensintervallet för μ ska minska till hälften, under förutsättning att s inte förändras? Hur stort stickprov behövs om vi vill halvera bredden och enbart förutsätter att $s \in I_\sigma$ enligt beräkningen av I_σ ovan? [4p]

Lösning:

(a) Vi har $n = 30$, $\alpha = 0.01$, $t_{\alpha/2}(n-1) = t_{0.005}(29) = 2.7564$, $s^2 = S_{xx}/(n-1) = 45.63/29 = 1.5734$ och $s = 1.2544$. Därmed får vi

$$I_\mu = \bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = 16.51 \pm 2.7564\frac{1.2544}{\sqrt{30}} = 16.51 \pm 0.6312.$$

Vidare har vi $\chi^2_{1-\alpha}(n-1) = \chi^2_{0.95}(29) = 17.708$, vilket ger det låga ensidiga konfidensintervallet

$$\begin{aligned} I_\sigma &= \left(0, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha}(n-1)}}\right) \\ &= \left(0, \sqrt{\frac{29 \cdot 1.5734}{17.708}}\right) \\ &= (0, 1.6052). \end{aligned}$$

(b) Eftersom intervalllängden delas med $\sqrt{n}$ behöver vi ha fyra gånger så många försök för att halvera intervallängden. Det krävs alltså ett stickprov med 120 försök.

Om vi enbart vet att $\sigma \in I_\sigma$ behöver vi utgå från intervallets högsta värde. Med $s \leq 1.605$ får vi ekvationen

$$\frac{1.605}{\sqrt{n}} = \frac{1}{2}\frac{1.2544}{\sqrt{30}},$$

vilket ger $n = 120 \cdot 1.605^2/1.5734 = 197$.